



## Reporting of radiographic methods in randomised controlled trials assessing structural outcomes in rheumatoid arthritis.

Gabriel Baron, Isabelle Boutron, Bruno Giraudeau, Philippe Ravaud

### ► To cite this version:

Gabriel Baron, Isabelle Boutron, Bruno Giraudeau, Philippe Ravaud. Reporting of radiographic methods in randomised controlled trials assessing structural outcomes in rheumatoid arthritis.. *Annals of the Rheumatic Diseases*, 2007, 66 (5), pp.651-7. 10.1136/ard.2006.063164 . inserm-00156552

**HAL Id: inserm-00156552**

**<https://www.hal.inserm.fr/inserm-00156552>**

Submitted on 22 Jun 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# **Title: Reporting of radiographic methods in randomized controlled trials assessing structural outcomes in rheumatoid arthritis**

**Authors:** Gabriel Baron<sup>1</sup>, Isabelle Boutron<sup>1</sup>, Bruno Giraudeau<sup>2</sup>, Philippe Ravaud<sup>1</sup>

<sup>1</sup>Gabriel Baron, MSc, Isabelle Boutron, MD, Philippe Ravaud, MD, PhD: INSERM U738, Paris, France ; Université Paris 7 Denis Diderot, UFR de Médecine, Paris, France ; AP-HP, Hôpital Bichat, Département d'Epidémiologie, Biostatistique et Recherche Clinique, Paris , France ;

<sup>2</sup>Bruno Giraudeau, PhD: INSERM CIC 202, Tours, France.

## **Address correspondence and reprint requests to**

Gabriel Baron, Département d'Epidémiologie Biostatistique et Recherche Clinique, INSERM U738, Groupe Hospitalier Bichat-Claude Bernard, 46 rue Henri Huchard, 75018 Paris, France.

Tel: +33 01 40 25 62 57

Fax: +33 01 40 25 67 73

E-mail: gabriel.baron@bch.aphp.fr

The Corresponding Author has the right to grant on behalf of all authors and does grant on behalf of all authors, an exclusive licence (or non exclusive for government employees) on a worldwide basis to the BMJ Publishing Group Ltd to permit this article (if accepted) to be published in ARD and any other BMJ PGL products and sublicences such use and exploit all subsidiary rights, as set out in our licence (<http://ARD.bmjournals.com/fora/licence.pdf>)

Word count: **2603**

## ABSTRACT

**Objective.** Because an increasing number of clinical trials evaluating disease modifying anti-rheumatic drugs in rheumatoid arthritis (RA) emphasize radiographic outcomes as a primary outcome, using a reproducible radiographic measure should be placed at a premium. We aimed to evaluate the reporting of radiographic methods in randomized trials assessing radiographic outcomes in RA.

**Methods.** We searched MEDLINE for randomized controlled trials assessing radiographic outcomes published between January 1994 and December 2005 in general medical and specialty journals with a high impact factor. One reader extracted data (radiographic acquisition, assessment and reproducibility) using a standardized form.

**Results.** A total of 46 reports were included in the analysis. The mean (SD) methodological quality scores on the Jadad scale (range 0-5) and the Delphi list (0-9) were 2.9 (1.2) and 6.4 (1.3), respectively. Use of a standardized procedure for the acquisition of the radiographs was reported in 2 articles (4.3%). Two reports (4.3%) indicated that the quality of the radiographs was evaluated. In 65.2% of the reports, 2 or more radiographic scores were used. Reporting of radiographic assessment was well detailed for number of readers (91.3%), information on readers (71.7%), blinding (91.4%) and how films were viewed (74.0%). The reproducibility of the reading was reported in 39.1% of the articles.

**Conclusion.** Our results highlight that the reporting of results of randomized controlled trials of radiographic outcomes in RA shows great variability in radiographic scores used and that reporting of radiographic methods could be improved upon, especially the acquisition procedure and the reproducibility of the reading.

**Key words:** rheumatoid arthritis, randomized controlled trials, reproducibility, radiographic acquisition, radiographic assessment

Rheumatoid arthritis (RA) is the most common chronic inflammatory joint disease and is responsible for symptomatic manifestations (e.g., functional status, pain) and structural damage (i.e., damage of the articular cartilage and bone) (1). The use of disease-modifying anti-rheumatic drugs has increased for RA (2). Assessing such treatments requires the measurement of structural outcomes in randomized controlled trials to demonstrate a reduction or a retardation of disease progression.

Radiography provides an objective measure of the extent of anatomical joint damage. It can be used to assess the severity of the structural destruction, to follow the course of the disease and to establish effects of treatment (3). This highly accepted technique is widely available and provides a permanent record of the structural image, allowing for comparison over time and rereading if necessary (4, 5). The assessment of radiographic outcomes for evaluating drug efficacy was recommended for the management of RA in controlled trials (6, 7) and the radiographic outcome is often used as a primary endpoint for assessing structural severity (8).

The reproducibility (i.e., the extent to which repeated measurements on the same subject yield the same results) is one of the prerequisites for a primary outcome (9-11). With radiography, measurements can be biased and their precision compromised by 2 well-identified sources of variability -- image acquisition and assessment -- which can be a serious limitation in its use. The image variability due to differences in acquisition processes is a major concern. Because many parameters can affect the appearance of the radiographs (i.e., positioning for radiographs, film exposure and resolution, reproducibility), standardization of the acquisition procedure is needed (12, 13). Similarly, radiographic assessment could be influenced by many known parameters (e.g., the scoring method, the number of readers, films grouped by patient or not, films chronologically scored or not) (14).

Evaluating the reproducibility of an outcome measure theoretically supposes a detailed reporting of the methods used to measure it. The CONSORT (Consolidated Standards of Reporting Trials) statement has recommended the reporting of methods used “to enhance the quality of measurements”, especially when considering the primary outcome (15), including how the outcome was measured and what steps were used to increase the reliability.

The purpose of this study was to evaluate the reporting of radiographic methods in randomized controlled trials assessing radiographic outcomes in RA. We investigated particularly radiographic acquisition, radiographic assessment and how the reproducibility was determined.

## **MATERIALS AND METHODS**

### **Search strategy for identification of studies**

The search method has been detailed elsewhere (16). Briefly, we searched MEDLINE and the Cochrane Central Register of Controlled Trials for randomized controlled trials assessing radiographic outcomes in RA published between January 1994 and December 2005 in general medical and specialty journals with a high impact factor (Appendix A). We chose these journals because a high impact factor is a good predictor of high methodological quality of journal articles (17) and because our goal was not to be exhaustive but, rather, to raise awareness of the reporting of radiographic assessment and acquisition. Potentially relevant articles were selected on the basis of the title, abstract and keywords by one reader (GB) using the following criteria: study population of adults aged 18 years or older, randomized controlled trial, and presence of at least 1 radiographic outcome evaluated by scheduled radiography. When duplicate publications of a trial existed, only the main publication was included (i.e., when a report of the same trial appeared twice or more, the report of the principal analysis planned by the protocol was selected). Reports of extended follow-up trials, analyses of multiple trials and subgroup analyses were excluded. When the abstract indicated that the article might be relevant, the entire paper was obtained.

### **Data extraction**

One author (GB) extracted all the data using a standardized form. From each article, data were obtained for the following:

(I) Characteristics of the included articles. Data were searched for number of centers, sample size, and methodological quality scores (Jadad scale (18) and the Delphi list (19)). The reviewer also determined whether the radiographic outcome was clearly defined as the primary endpoint.

(II) Radiographic acquisition. The reviewer extracted data on regions assessed (e.g., hand, feet), number of radiographic sessions, time between baseline session and the last session, and time between 2 radiographic sessions. The reviewer checked whether methods used to take the radiographs were detailed: radiographic view (e.g., dorsovolat, anteroposterior), use of a standardized procedure to improve positioning, description of the image quality (film exposure and use of resolution films), and use of digitized image or not.

(III) Radiographic assessment. The reviewer collected data on which score was used for primary and secondary radiographic analysis and determined whether information was included on the reader's experience (e.g., it was explicitly stated that radiographs were read by a well-known expert), background (e.g., radiologist) and identity (e.g., initials reported); and whether there were multiple observers (i.e., methods of consensus to obtain a single score described when appropriate). The reviewer also checked whether the assessment of the radiographs was blinded (i.e., to treatment assignment or patient identity and clinical data). He also determined how films were read (paired reading with chronological or random order).

(IV) Reproducibility. The reviewer recorded whether intra- and inter-reader reproducibility was evaluated and how.

## **Analysis**

Categorical variables were described with frequencies and percentages and quantitative variables with means (SD) or medians [minimum, maximum]. Two ancillary analyses were

also performed. The first describes characteristics of the radiographic assessment in reports published before 2000 versus reports published in 2000 and later and the second compares reports with radiographic outcome defined as a primary endpoint and reports with a nonradiographic primary outcome. All data analyses involved use of SAS for Windows, Release 9.1 (SAS Institute, Cary, NC).



## **RESULTS**

### **Characteristics of the included articles**

The search strategy generated 1004 citations. Forty-six studies were relevant according to the title, abstract, and complete retrieval of the article (figure 1). Detailed references of the selected articles are listed in appendix B.

Of the reports, 40 (87.0%) concerned multicenter studies. The median sample size of the studies was 183 [20, 1446] patients. The mean (SD) methodological quality scores were 2.8 (1.0) for the Jadad scale and 6.5 (1.3) for the Delphi list. The radiographic outcome was clearly defined as the primary endpoint in 26 reports (56.5%).

### **Radiographic acquisition**

Both joints of hand and feet were the most frequently assessed areas (75.0%) (table 1). In many reports (n=20, 43.5%), 2 radiographs (included baseline radiographs) were obtained (table 1). The median time between the baseline session and the last session was 12 months [5.5, 48 months] and the median time between 2 sessions was 53 weeks [24, 104 weeks].

The radiographic view (e.g., dorsovolar) was described in 11 reports (23.9%). Use of a standardized radiographic procedure for optimal positioning was reported in 2 articles (4.3%). The X-ray films were described as digitized in 3 reports (6.5%). The assessment of radiographic quality was reported in only 2 articles (4.3%) and details on film resolution (e.g., single emulsion film) were provided in 2 (4.3%) reports.

### **Radiographic assessment**

### ***Radiographic score***

In 65.2% of the reports, 2 or more radiographic scores were used (table 2). Among scores combining erosions and joint space narrowing, the Sharp score and the Larsen score (including their modified versions) were the most represented methods (47.8% and 34.7%, respectively).

### ***Readers***

Information on reader(s) (i.e., readers' experience in evaluating radiographs, readers' background or identity) were described in 33 articles (71.7%). Readers' experience, background and identity were reported in 17 (37.0%), 25 (54.3%), and 23 (50.0%) reports, respectively (table 3).

In 4 reports (8.7%), the number of readers was impossible to determine. Radiographs were read by at least 2 readers in 30 reports (65.2%) and by a single reader in 12 (26.1%).

Of the reports describing at least 2 readers, 22 described readers reading all the radiographs (61.1%) and 1 all the readers did not read all the radiographs (3.3%), and in 7 the reading was unclear (23.3%). When all the radiographs were read (n=22), a consensus method to produce a single radiographic score was described in 8 reports (consensus with the same readers for all disagreements [n=5], consensus for only important disagreements [n=2], or use of another method of measurement [n=1]); in 10 reports, the radiographic score used came from the mean score of the readers and from the lower score of the readers in 2 reports (in the 2 remaining reports, it was unclear how a single score was obtained).

### ***Blinding***

Masked assessment of radiographs was described in 43 (93.5%) reports (Table 3). Blinding was described to concern treatment assignment or patient identity (n=37, 86.0%) and clinical data (n=11, 23.9%); 4 articles reported that the assessor was blinded, with no other details given.

### ***Reading session***

In 12 articles (26.1%), no details were provided about the reading of the radiographs. When reading was described (73.9%), paired reading in random order and paired reading in chronological order were used in 18 (52.9%) and 15 (44.1%) articles, respectively. One article (2.9%) used both random and chronological paired reading.

### ***Reproducibility of the reading***

Reproducibility of the reading was described in 18 articles (39.1%). Intrareader reproducibility was reported in 9 (19.6) articles and inter-reader reproducibility in 13 (28.3) (4 articles [8.7%] reported both intra- and inter-reader reproducibility). Five articles (10.8%) reported 2 reproducibility measures. Use of the intraclass correlation coefficient (n=8) and coefficient of correlation (n=7) and reporting the smallest detectable difference (n=6) were the most common methods used to assess reproducibility. Other methods were use of median with minimum and maximum (n=1).

### ***Ancillary analysis***

When comparing reports with radiographic outcome defined as a primary endpoint versus other reports, we only found a major difference in the reporting of reproducibility (53.8% versus 20.0%) (table 3).

No difference was found in the reporting of radiographic assessment between reports published before 2000 and reports published in 2000 and later (table 3).

## DISCUSSION

Because an increasing number of clinical trials in RA emphasize radiographic outcome as a primary outcome, investigating the reporting of the radiographic methods is important. Our results show a great variability of the radiographic scores used and that radiographic assessment was better described than radiographic acquisition. Measures of reproducibility were reported in almost 40% of the assessed reports.

Reducing measurement error is an important objective, so use of a reproducible radiographic measure should be placed at a premium. Reporting reproducibility measures and describing methods of measurement are an essential step to confirm the validity of a measure (15). Controlling the reproducibility of a radiographic measurement concerns 2 steps: acquisition and assessment.

Although variability of acquisition can lead to measurement error (12, 13), only a few articles described radiographic acquisition in detail. The standardization of the acquisition is facilitated by centralized acquisition and entails training in radiographic acquisition and similar conditions of assessment (e.g., positioning for radiography (20)). Even if centralization of the acquisition is not always possible (e.g., because of multicenter studies or financial reasons), the training of radiologists could be centralized to decrease the variability. Radiographic assessment also depends on the technical quality of the radiographs, such as film exposure or resolution (21). So, evaluation of radiographic quality could be a control element that allows for limiting the number of nonassessable radiographs.

When designing a trial, the choice of a radiographic score for assessing structural destruction can directly influence reproducibility (21, 22). When focusing on radiographic scores used in assessed reports, we found no consensus in assessment scores used, probably

because the radiographic score use has evolved over the years (21). However, in the last few years, the modified Sharp and Larsen scores seem to have been preferentially used (23).

When considering other radiographic assessment characteristics, we noted that blindness was well respected and described in our study. Similarly, recommendations on reporting the number of observers and describing how to obtain a single measurement are followed (e.g., mean radiographic score from 2 assessors to decrease measurement error (24)). Specific training, as a calibration exercise, could be requested to come to an agreement when there are more than 1 reader. Some other parameters could be better described. For instance, experienced readers have been shown to have better agreement (22). A number of reports did not state in what order radiographs were assessed, even though this parameter can influence measurements (23, 25).

Almost 40% of the articles described the reproducibility of reading, even though reporting agreement between observers is essential to assessing the quality of observations (14, 26). If reproducibility was low or not assessed, then the use of radiographic outcomes as a primary endpoint may be questionable. Most of the methods reported, such use of as intraclass correlation coefficient or smallest detectable difference, were adapted for evaluating reproducibility. However, the use of the correlation coefficient as a measure of agreement was found in a surprising number of articles. Correlation coefficients measure the strength of an association, not the concordance, and thus should be avoided in this indication (27).

Van der Heijde et al. give recommendations for improving the reporting of radiographic methods in studies of radiographic outcomes (14). Because the optimal interpretation of a radiograph also depends on conditions of acquisition, we suggest that the radiographic acquisition should be more detailed or referenced in reports. Such a description could entail whether a standardized protocol was followed, whether the technicians were

educated, whether the radiographs were digitized, whether the quality of films was assessed, which view was chosen and what kind of film was used.

Our study is limited in that our search was restricted to articles published in high-impact-factor journals. However, articles in low-impact-factor journals may have had the same or lower methodological quality. Second, we pooled phase II and III studies with those of a more epidemiological nature, even if they were also randomized controlled trials. The degree of conformance with regulatory requirements and, consequently, required details and rigor are probably more important for reporting phase II and phase III trials. Third, the radiographic outcome was not always defined as a primary endpoint in our selected studies. We could have presumed a more detailed description of the conditions of measurement if we considered only reports in which the radiographic outcome was the primary endpoint, but results were quite similar. We also decided to include all the studies in all the fields evaluated because the primary outcome has been shown to differ between protocols and reports of studies. In fact, Chan et al. demonstrated that primary outcome differed between protocol and publication in 40% to 62% of trials (28, 29). Finally, some discrepancies may exist between the real methods used and methods reported. Some deficiencies may appear simply because of poor reporting, which does not necessarily mean that the methods were not applied (30, 31). In fact, some details may not have been reported because the authors regarded them to be standard and not necessary to mention (e.g., radiographic view or details about film resolution). Similarly, authors are often forced by referees or editors to abbreviate the report, and so important information such as details on acquisition technique are removed from manuscripts. However, online appendices, supplemental information, or longer versions of a paper could be provided. For example, the U.S. Food and Drug Administration provided in its Website most of the required details of the radiographic methods used in reporting results of a trial.

In conclusion, our results have highlighted that the reporting of results of randomized controlled trials of radiographic outcomes measured by scheduled radiography in RA showed a great variability of radiographic scores used and that reporting of radiographic methods could be improved upon, especially the acquisition procedure and the reproducibility. Investigators are encouraged to follow guidelines on reporting radiographic data in randomized controlled trials in RA.



## REFERENCES

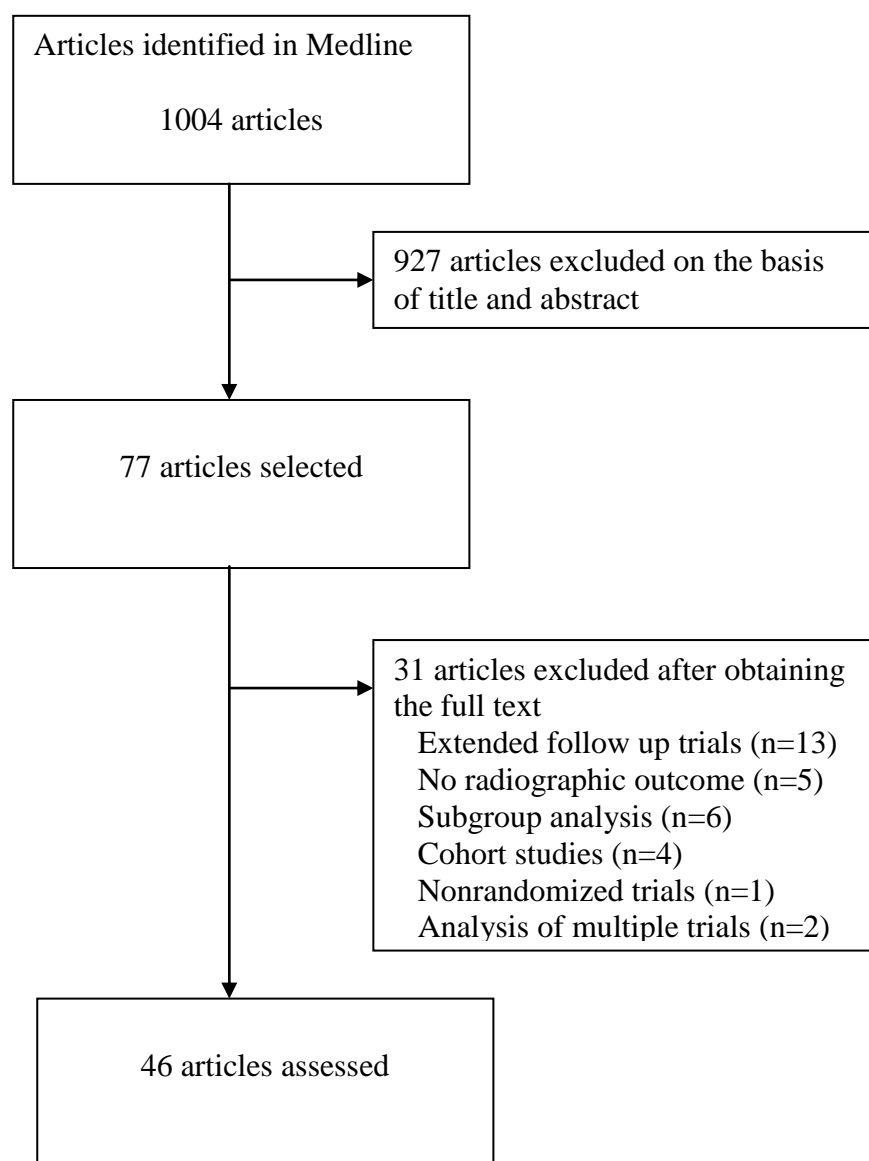
1. Symmons DP. Disease assessment indices: activity, damage and severity. *Baillieres Clin Rheumatol* 1995;9:267-85.
2. Lee DM, Weinblatt ME. Rheumatoid arthritis. *Lancet* 2001;358:903-11.
3. Soubrier M, Dougados M. How to assess early rheumatoid arthritis in daily clinical practice. *Best Pract Res Clin Rheumatol* 2005;19:73-89.
4. van der Heijde DM. Plain X-rays in rheumatoid arthritis: overview of scoring methods, their reliability and applicability. *Baillieres Clin Rheumatol* 1996;10:435-53.
5. Pincus T, Sokka T. Quantitative measures for assessing rheumatoid arthritis in clinical trials and clinical care. *Best Pract Res Clin Rheumatol* 2003;17:753-81.
6. Felson DT, Anderson JJ, Boers M, Bombardier C, Chernoff M, Fried B, et al. The American College of Rheumatology preliminary core set of disease activity measures for rheumatoid arthritis clinical trials. The Committee on Outcome Measures in Rheumatoid Arthritis Clinical Trials. *Arthritis Rheum* 1993;36:729-40.
7. Recommendations for the registration of drugs used in the treatment of rheumatoid arthritis. Group for the Respect of Ethics and Excellence in Science (GREES): rheumatoid arthritis section. *Br J Rheumatol* 1998;37:211-5.
8. Boers M, van der Heijde DM. Prevention or retardation of joint damage in rheumatoid arthritis: issues of definition, evaluation and interpretation of plain radiographs. *Drugs* 2002;62:1717-24.
9. Fleiss JL. Reliability of measurement. In: *The design and analysis of clinical experiments*. New York: Wiley; 1986. p. 1-32.
10. Giraudeau B, Ravaud P, Chastang C. Importance of reproducibility in responsiveness issues. *Biom J* 1998;40:685-701.
11. Vangeneugden T, Laenen A, Geys H, Renard D, Molenberghs G. Applying concepts of generalizability theory on clinical trial data to investigate sources of variation and their impact on reliability. *Biometrics* 2005;61:295-304.
12. Brower AC. Use of the radiograph to measure the course of rheumatoid arthritis. The gold standard versus fool's gold. *Arthritis Rheum* 1990;33:316-24.
13. Weissman BN. Imaging techniques in rheumatoid arthritis. *J Rheumatol Suppl* 1994;42:14-9.

14. van der Heijde D, Simon L, Smolen J, Strand V, Sharp J, Boers M, et al. How to report radiographic data in randomized clinical trials in rheumatoid arthritis: guidelines from a roundtable discussion. *Arthritis Rheum* 2002;47:215-8.
15. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.
16. Baron G, Boutron I, Giraudeau B, Ravaud P. Violation of the intent-to-treat principle and rate of missing data in superiority trials assessing structural outcomes in rheumatic diseases. *Arthritis Rheum* 2005;52:1858-1865.
17. Lee KP, Schotland M, Bacchetti P, Bero LA. Association of journal quality indicators with methodological quality of clinical research articles. *Jama* 2002;287:2805-8.
18. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17:1-12.
19. Verhagen AP, de Vet HC, de Bie RA, Kessels AG, Boers M, Bouter LM, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol* 1998;51:1235-41.
20. van der Heijde DM. Radiographic imaging: the 'gold standard' for assessment of disease progression in rheumatoid arthritis. *Rheumatology (Oxford)* 2000;39 Suppl 1:9-16.
21. Boini S, Guillemin F. Radiographic scoring methods as outcome measures in rheumatoid arthritis: properties and advantages. *Ann Rheum Dis* 2001;60:817-27.
22. Sharp JT, Wolfe F, Lassere M, Boers M, Van Der Heijde D, Larsen A, et al. Variability of precision in scoring radiographic abnormalities in rheumatoid arthritis by experienced readers. *J Rheumatol* 2004;31:1062-72.
23. Ory PA. Interpreting radiographic data in rheumatoid arthritis. *Ann Rheum Dis* 2003;62:597-604.
24. Van Der Heijde D. Structural damage in rheumatoid arthritis as visualized through radiographs. *Arthritis Res* 2002;4 Suppl 2:S29-33.
25. van Der Heijde D, Boonen A, Boers M, Kostense P, van Der Linden S. Reading radiographs in chronological order, in pairs or as single films has important implications for the discriminative power of rheumatoid arthritis clinical trials. *Rheumatology (Oxford)* 1999;38:1213-20.
26. van der Heijde D, Boers M, Lassere M. Methodological issues in radiographic scoring methods in rheumatoid arthritis. *J Rheumatol* 1999;26:726-30.

27. Lin LI. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* 1989;45:255-68.
28. Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *Jama* 2004;291:2457-65.
29. Chan AW, Krleza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *Cmaj* 2004;171:735-40.
30. Soares HP, Daniels S, Kumar A, Clarke M, Scott C, Swann S, et al. Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *Bmj* 2004;328:22-4.
31. Hill CL, LaValley MP, Felson DT. Discrepancy between published report and actual conduct of randomized clinical trials. *J Clin Epidemiol* 2002;55:783-6.
32. Sharp JT, Lidsky MD, Collins LC, Moreland J. Methods of scoring the progression of radiologic changes in rheumatoid arthritis. Correlation of radiologic, clinical and laboratory abnormalities. *Arthritis Rheum* 1971;14:706-20.
33. Sharp JT, Young DY, Bluhm GB, Brook A, Brower AC, Corbett M, et al. How many joints in the hands and wrists should be included in a score of radiologic abnormalities used to assess rheumatoid arthritis? *Arthritis Rheum* 1985;28:1326-35.
34. van der Heijde D, Dankert T, Nieman F, Rau R, Boers M. Reliability and sensitivity to change of a simplification of the Sharp/van der Heijde radiological assessment in rheumatoid arthritis. *Rheumatology (Oxford)* 1999;38:941-7.
35. van der Heijde DM, van Riel PL, Nuvér-Zwart IH, Gribnau FW, van de Putte LB. Effects of hydroxychloroquine and sulphasalazine on progression of joint damage in rheumatoid arthritis. *Lancet* 1989;1:1036-8.
36. Larsen A, Dale K, Eek M. Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films. *Acta Radiol Diagn (Stockh)* 1977;18:481-91.
37. Larsen A, Thoen J. Hand radiography of 200 patients with rheumatoid arthritis repeated after an interval of one year. *Scand J Rheumatol* 1987;16:395-401.
38. Larsen A. How to apply Larsen score in evaluating radiographs of rheumatoid arthritis in long-term studies. *J Rheumatol* 1995;22:1974-5.
39. Rau R, Wassenberg S, Herborn G, Stucki G, Gebler A. A new method of scoring radiographic change in rheumatoid arthritis. *J Rheumatol* 1998;25:2094-107.

40. Genant HK, Jiang Y, Peterfy C, Lu Y, Redei J, Countryman PJ. Assessment of rheumatoid arthritis using a modified scoring method on digitized and original radiographs. *Arthritis Rheum* 1998;41:1583-90.

**Figure 1.** Process of screening articles on randomized controlled trials of radiographic outcomes in rheumatoid arthritis, for inclusion in the present analysis.



**Table 1.** Generic data on radiographs described in reports of randomized controlled trials of radiographic outcomes

<b>Generic data</b>	<b>N</b>	<b>%</b>
<b>Assessed areas</b>	44/46	95.7
Hand and feet	33/44	75.0
Hand	9/44	20.4
Hand and ankle	1/44	02.3
Shoulders, elbows, hips, knees, ankles and subtalar joints	1/44	02.3
<b>Number of radiographic sessions</b>	46/46	100.0
Baseline and 1 radiograph	20/46	43.5
Baseline and 2 radiographs	18/46	39.1
Baseline and 3 or more radiographs	8/46	17.4
<b>Radiographic view reported</b>	11/46	23.9
Posteroanterior (e.g., dorsovolar for hands)	4/11	36.4
Anteroposterior	5/11	45.5
Both	2/11	18.1

**Table 2.** Radiographic outcomes in randomized controlled trials in rheumatoid arthritis

<b>Radiographic outcomes</b>	<b>N=46</b>	<b>%</b>
Sharp score (original (32), modified by Sharp (33) or by van der Heijde's (34, 35))	22	47.8
Erosion score and joint space narrowing score reported with composite score	16	34.8
Larsen score (original (36, 37) or modified version (38))	18	39.0
Number of eroded joints counts	9	19.6
Erosion score and joint space narrowing score reported without composite score	5	10.9
Number of erosions	3	06.5
Raitingen score (39)	2	04.3
Cumulative number of joints free of erosions at baseline in which at least one erosion developed during follow-up	2	04.3
Genant scoring method (40)	1	02.2
Others	7	15.2

**Table 3.** Reporting of radiographic assessment in reports of randomized controlled trials with radiographic outcomes in rheumatoid arthritis \*

Reporting of	All N=46	Reports with radiographic outcome as primary endpoint N=26	Reports with radiographic outcome not as primary endpoint N=20	Reports published before 2000 N=17	Reports published in 2000 and later N=29
<b>Reader's information</b>	71.7	73.1	70.0	70.6	72.4
Experience	36.9	34.6	40.0	35.3	64.7
Background	54.4	57.7	50.0	52.9	55.2
Identity	50.0	50.0	50.0	52.9	48.3
<b>Number of readers</b>	91.3	96.2	85.0	94.1	89.7
1	26.1	19.2	35.0	29.4	24.2
>1	65.2	76.9	50.0	64.7	65.5
Not reported	08.7	03.9	15.0	05.9	10.3
<b>Blinding</b>	93.5	96.2	90.0	94.1	93.1
To treatment or patient identity	86.0	92.0	77.8	87.5	85.2
To clinical data	23.9	26.9	20.0	17.6	27.6
<b>How films were viewed</b>	73.9	80.7	65.0	70.6	75.9
<b>Reproducibility</b>	39.1	53.8	20.0	35.3	41.4

\* All data are expressed as percentages.



**Appendix A.** Journals included in the search strategy for randomized controlled trials with radiographic outcomes in rheumatoid arthritis

10 general and internal medicine journals (*New England Journal of Medicine*, *Journal of the American Medical Association*, *The Lancet*, *Annals of Internal Medicine*, *Annual Review of Medicine*, *Archives of Internal Medicine*, *British Medical Journal*, *American Journal of Medicine*, *Medicine*, and *Proceedings of the Association of American Physicians*);

6 rheumatologic journals (*Arthritis and Rheumatism*, *Seminars in Arthritis and Rheumatism*, *Annals of Rheumatic Diseases*, *Rheumatology [Oxford, England]*, *Journal of Rheumatology*, and *Arthritis Care and Research*);

6 orthopedic journals (*Osteoarthritis and Cartilage/OARS*, *Arthroscopy*, *Journal of Orthopaedic Research: Official Publication of the Orthopaedic Research Society*, *Journal of Bone and Joint Surgery*, *American Volume*, *Spine*, and *Gait & Posture*);

6 rehabilitation journals (*Archives of Physical Medicine and Rehabilitation*, *Supportive Care in Cancer: Official Journal of the Multinational Association of Supportive Care in Cancer*, *Journal of Electromyography and Kinesiology: Official Journal of the International Society of Electrophysiological Kinesiology*, *Physical Therapy*, *Journal of Rehabilitation Research and Development*, and *Scandinavian Journal of Rehabilitation Medicine*).

## **Appendix B.** Detailed references of selected articles

1. Bathon JM, Martin RW, Fleischmann RM, Tesser JR, Schiff MH, Keystone EC, et al. A comparison of etanercept and methotrexate in patients with early rheumatoid arthritis. *N Engl J Med* 2000;343:1586-93.
2. Bluhm GB, Sharp JT, Tilley BC, Alarcon GS, Cooper SM, Pillemer SR, et al. Radiographic results from the Minocycline in Rheumatoid Arthritis (MIRA) Trial. *J Rheumatol* 1997;24:1295-302.
3. Boers M, Verhoeven AC, Markusse HM, van de Laar MA, Westhovens R, van Denderen JC, et al. Randomised comparison of combined step-down prednisolone, methotrexate and sulphasalazine with sulphasalazine alone in early rheumatoid arthritis. *Lancet* 1997;350:309-18.
4. Bresnihan B, Alvaro-Gracia JM, Cobby M, Doherty M, Domljan Z, Emery P, et al. Treatment of rheumatoid arthritis with recombinant human interleukin-1 receptor antagonist. *Arthritis Rheum* 1998;41:2196-204.
5. Capell HA, Madhok R, Hunter JA, Porter D, Morrison E, Larkin J, et al. Lack of radiological and clinical benefit over two years of low dose prednisolone for rheumatoid arthritis: results of a randomised controlled trial. *Ann Rheum Dis* 2004;63:797-803.
6. Choy EH, Kingsley GH, Khoshaba B, Pipitone N, Scott DL. A two year randomised controlled trial of intramuscular depot steroids in patients with established rheumatoid arthritis who have shown an incomplete response to disease modifying antirheumatic drugs. *Ann Rheum Dis* 2005;64:1288-93.
7. de Jong Z, Munneke M, Zwinderman AH, Kroon HM, Jansen A, Runday KH, et al. Is a long-term high-intensity exercise program effective and safe in patients with rheumatoid arthritis? Results of a randomized controlled trial. *Arthritis Rheum* 2003;48:2415-24.
8. D'Elia HF, Larsen A, Mattsson LA, Waltbrand E, Kvist G, Mellstrom D, et al. Influence of hormone replacement therapy on disease progression and bone mineral density in rheumatoid arthritis. *J Rheumatol* 2003;30:1456-63.
9. Dougados M, Combe B, Cantagrel A, Goupille P, Olive P, Schattenkirchner M, et al. Combination therapy in early rheumatoid arthritis: a randomised, controlled, double blind 52 week clinical trial of sulphasalazine and methotrexate compared with the single components. *Ann Rheum Dis* 1999;58:220-5.

10. Eggelmeijer F, Papapoulos SE, van Paassen HC, Dijkmans BA, Valkema R, Westedt ML, et al. Increased bone mass with pamidronate treatment in rheumatoid arthritis. Results of a three-year randomized, double-blind trial. *Arthritis Rheum* 1996;39:396-402.
11. Emery P, Breedveld FC, Lemmel EM, Kaltwasser JP, Dawes PT, Gomor B, et al. A comparison of the efficacy and safety of leflunomide and methotrexate for the treatment of rheumatoid arthritis. *Rheumatology (Oxford)* 2000;39:655-65.
12. Forre O. Radiologic evidence of disease modification in rheumatoid arthritis patients treated with cyclosporine. Results of a 48-week multicenter study comparing low-dose cyclosporine with placebo. Norwegian Arthritis Study Group. *Arthritis Rheum* 1994;37:1506-12.
13. Gerards AH, Landewe RB, Prins AP, Bruijn GA, Goei The HS, Laan RF, et al. Cyclosporin A monotherapy versus cyclosporin A and methotrexate combination therapy in patients with early rheumatoid arthritis: a double blind randomised placebo controlled trial. *Ann Rheum Dis* 2003;62:291-6.
14. Goekoop-Ruiterman YP, de Vries-Bouwstra JK, Allaart CF, van Zeben D, Kerstens PJ, Hazes JM, et al. Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): a randomized, controlled trial. *Arthritis Rheum* 2005;52:3381-90.
15. Hafstrom I, Ringertz B, Spangberg A, von Zweigbergk L, Brannemark S, Nylander I, et al. A vegan diet free of gluten improves the signs and symptoms of rheumatoid arthritis: the effects on arthritis correlate with a reduction in antibodies to food antigens. *Rheumatology (Oxford)* 2001;40:1175-9.
16. Hakkinen A, Sokka T, Kotaniemi A, Hannonen P. A randomized two-year study of the effects of dynamic strength training on muscle strength, disease activity, functional capacity, and bone mineral density in early rheumatoid arthritis. *Arthritis Rheum* 2001;44:515-22.
17. Hansen M, Podenphant J, Florescu A, Stoltenberg M, Borch A, Kluger E, et al. A randomised trial of differentiated prednisolone treatment in active rheumatoid arthritis. Clinical benefits and skeletal side effects. *Ann Rheum Dis* 1999;58:713-8.
18. Hasegawa J, Nagashima M, Yamamoto M, Nishijima T, Katsumata S, Yoshino S. Bone resorption and inflammatory inhibition efficacy of intermittent cyclical etidronate therapy in rheumatoid arthritis. *J Rheumatol* 2003;30:474-9.
19. Helliwell PS, O'Hara M, Holdsworth J, Hesselden A, King T, Evans P. A 12-month randomized controlled trial of patient education on radiographic changes and quality of life in early rheumatoid arthritis. *Rheumatology (Oxford)* 1999;38:303-8.

20. Jiang Y, Genant HK, Watt I, Cobby M, Bresnihan B, Aitchison R, et al. A multicenter, double-blind, dose-ranging, randomized, placebo-controlled study of recombinant human interleukin-1 receptor antagonist in patients with rheumatoid arthritis: radiologic progression and correlation of Genant and Larsen scores. *Arthritis Rheum* 2000;43:1001-9.
21. Keystone EC, Kavanaugh AF, Sharp JT, Tannenbaum H, Hua Y, Teoh LS, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. *Arthritis Rheum* 2004;50:1400-11.
22. Kirwan JR. The effect of glucocorticoids on joint destruction in rheumatoid arthritis. The Arthritis and Rheumatism Council Low-Dose Glucocorticoid Study Group. *N Engl J Med* 1995;333:142-6.
23. Klareskog L, van der Heijde D, de Jager JP, Gough A, Kalden J, Malaise M, et al. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. *Lancet* 2004;363:675-81.
24. Lipsky PE, van der Heijde DM, St Clair EW, Furst DE, Breedveld FC, Kalden JR, et al. Infliximab and methotrexate in the treatment of rheumatoid arthritis. Anti-Tumor Necrosis Factor Trial in Rheumatoid Arthritis with Concomitant Therapy Study Group. *N Engl J Med* 2000;343:1594-602.
25. Marchesoni A, Battafarano N, Arreghini M, Panni B, Gallazzi M, Tosi S. Radiographic progression in early rheumatoid arthritis: a 12-month randomized controlled study comparing the combination of cyclosporin and methotrexate with methotrexate alone. *Rheumatology (Oxford)* 2003;42:1545-9.
26. Marchesoni A, Battafarano N, Arreghini M, Pellerito R, Cagnoli M, Prudente P, et al. Step-down approach using either cyclosporin A or methotrexate as maintenance therapy in early rheumatoid arthritis. *Arthritis Rheum* 2002;47:59-66.
27. Mottonen T, Hannonen P, Leirisalo-Repo M, Nissila M, Kautiainen H, Korpela M, et al. Comparison of combination therapy with single-drug therapy in early rheumatoid arthritis: a randomised trial. FIN-RACo trial group. *Lancet* 1999;353:1568-73.
28. Neustadt DH. Double blind evaluation of the long-term effects of etodolac versus ibuprofen in patients with rheumatoid arthritis. *J Rheumatol Suppl* 1997;47:17-22.

29. Pasero G, Priolo F, Marubini E, Fantini F, Ferraccioli G, Magaro M, et al. Slow progression of joint damage in early rheumatoid arthritis treated with cyclosporin A. *Arthritis Rheum* 1996;39:1006-15.
30. Proudman SM, Conaghan PG, Richardson C, Griffiths B, Green MJ, McGonagle D, et al. Treatment of poor-prognosis early rheumatoid arthritis. A randomized study of treatment with methotrexate, cyclosporin A, and intraarticular corticosteroids compared with sulfasalazine alone. *Arthritis Rheum* 2000;43:1809-19.
31. Quinn MA, Conaghan PG, O'Connor PJ, Karim Z, Greenstein A, Brown A, et al. Very early treatment with infliximab in addition to methotrexate in early, poor-prognosis rheumatoid arthritis reduces magnetic resonance imaging evidence of synovitis and damage, with sustained benefit after infliximab withdrawal: results from a twelve-month randomized, double-blind, placebo-controlled trial. *Arthritis Rheum* 2005;52:27-35.
32. Rau R, Herborn G, Menninger H, Sangha O. Radiographic outcome after three years of patients with early erosive rheumatoid arthritis treated with intramuscular methotrexate or parenteral gold. Extension of a one-year double-blind study in 174 patients. *Rheumatology (Oxford)* 2002;41:196-204.
33. Smolen JS, Kalden JR, Scott DL, Rozman B, Kvien TK, Larsen A, et al. Efficacy and safety of leflunomide compared with placebo and sulphasalazine in active rheumatoid arthritis: a double-blind, randomised, multicentre trial. European Leflunomide Study Group. *Lancet* 1999;353:259-66.
34. St Clair EW, van der Heijde DM, Smolen JS, Maini RN, Bathon JM, Emery P, et al. Combination of infliximab and methotrexate therapy for early rheumatoid arthritis: a randomized, controlled trial. *Arthritis Rheum* 2004;50:3432-43.
35. Strand V, Cohen S, Schiff M, Weaver A, Fleischmann R, Cannon G, et al. Treatment of active rheumatoid arthritis with leflunomide compared with placebo and methotrexate. Leflunomide Rheumatoid Arthritis Investigators Group. *Arch Intern Med* 1999;159:2542-50.
36. Svensson B, Boonen A, Albertsson K, van der Heijde D, Keller C, Hafstrom I. Low-dose prednisolone in addition to the initial disease-modifying antirheumatic drug in patients with early active rheumatoid arthritis reduces joint destruction and increases the remission rate: a two-year randomized trial. *Arthritis Rheum* 2005;52:3360-70.
37. Taylor PC, Steuer A, Gruber J, Cosgrove DO, Blomley MJ, Marsters PA, et al. Comparison of ultrasonographic assessment of synovitis and joint vascularity with radiographic evaluation in a randomized, placebo-controlled study of infliximab therapy in early rheumatoid arthritis. *Arthritis Rheum* 2004;50:1107-16.

38. Valleala H, Laasonen L, Koivula MK, Mandelin J, Friman C, Risteli J, et al. Two year randomized controlled trial of etidronate in rheumatoid arthritis: changes in serum aminoterminal telopeptides correlate with radiographic progression of disease. *J Rheumatol* 2003;30:468-73.
39. van der Heide A, Jacobs JW, Bijlsma JW, Heurkens AH, van Booma-Frankfort C, van der Veen MJ, et al. The effectiveness of early treatment with "second-line" antirheumatic drugs. A randomized, controlled trial. *Ann Intern Med* 1996;124:699-707.
40. van der Laan W, Molenaar E, Roday K, Verheijen J, Breedveld F, Greenwald R, et al. Lack of effect of doxycycline on disease activity and joint damage in patients with rheumatoid arthritis. A double blind, placebo controlled trial. *J Rheumatol* 2001;28:1967-74.
41. van Everdingen AA, Jacobs JW, Siewerts Van Reesema DR, Bijlsma JW. Low-dose prednisone therapy for patients with early active rheumatoid arthritis: clinical efficacy, disease-modifying properties, and side effects: a randomized, double-blind, placebo-controlled clinical trial. *Ann Intern Med* 2002;136:1-12.
42. van Holten J, Pavelka K, Vencovsky J, Stahl H, Rozman B, Genovese M, et al. A multicentre, randomised, double blind, placebo controlled phase II study of subcutaneous interferon beta-1a in the treatment of patients with active rheumatoid arthritis. *Ann Rheum Dis* 2005;64:64-9.
43. van Jaarsveld CH, Jacobs JW, van der Veen MJ, Blaauw AA, Kruize AA, Hofman DM, et al. Aggressive treatment in early rheumatoid arthritis: a randomised controlled trial. On behalf of the Rheumatic Research Foundation Utrecht, The Netherlands. *Ann Rheum Dis* 2000;59:468-77.
44. van Schaardenburg D, Valkema R, Dijkmans BA, Papapoulos S, Zwinderman AH, Han KH, et al. Prednisone treatment of elderly-onset rheumatoid arthritis. Disease activity and bone mass in comparison with chloroquine treatment. *Arthritis Rheum* 1995;38:334-42.
45. Wassenberg S, Rau R, Steinfeld P, Zeidler H. Very low-dose prednisolone in early rheumatoid arthritis retards radiographic progression over two years: a multicenter, double-blind, placebo-controlled trial. *Arthritis Rheum* 2005;52:3371-80.
46. Willkens RF, Sharp JT, Stablein D, Marks C, Wortmann R. Comparison of azathioprine, methotrexate, and the combination of the two in the treatment of rheumatoid arthritis. A forty-eight-week controlled clinical trial with radiologic outcome assessment. *Arthritis Rheum* 1995;38:1799-806.